

2X increase

in throughput on some large language models by moving to Intel® Gaudi® 2.<sup>1</sup>

“Our participation in Intel Liftoff was critical for getting us where we are today. It allowed us to explore the art of the possible with Intel Gaudi 2 processors and demonstrate early on that our system could run effectively on Intel Developer Cloud at the speed and scale we needed.”

**Daniel Whitenack,**  
Founder, Prediction Guard

# Prediction Guard De-Risks Large Language Model Applications at Scale

Prediction Guard, a startup company operating at the forefront of AI integration, wants to help organizations unleash that potential faster. The company’s clients want to use Large Language Models (LLM) in generative AI (GenAI) applications such as information extraction, customer service and analytics, marketing campaign planning, financial reporting, and “co-pilot” solutions. While LLMs are proving their utility across a wide range of GenAI applications, they can bring risks when it comes to reliability, variability, and security. Prediction Guard is working to reduce those risks while using Intel® Developer Cloud and Intel® Gaudi® 2 processors to run more efficiently and effectively as it helps organizations integrate LLMs into their AI-enabled business applications.

Products and Solutions  
[Intel® Developer Cloud](#)  
[Intel® Gaudi® 2 Processors](#)  
[oneAPI](#)

Industry  
Software  
Development

Organization Size  
2-10

Country  
United States

Learn more  
[Case Study](#)